ECE 509: Convex Optimization **Rutgers: Spring 2024**

Homework #3

February 22, 2024

Name: Ravi Raghavan *Extension: No*

Problem 1. Suppose $f : \mathbb{R} \to \mathbb{R}$ is convex, and $a, b \in dom f$ with $a < b$

(a) Show that

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b) \tag{1}$$

for all $x \in [a, b]$.

Solution. Let us have $\theta \in \mathbb{R}$ where $0 \leq \theta \leq 1$. Since $x \in [a, b]$, we can define x as $\theta a + (1-\theta)b$

Based on the definition of a convex function, we know that $f(\theta a + (1-\theta)b) \leq \theta f(a) + (1-\theta)f(b)$

$f(x) \leq \theta f(a) + (1-\theta)f(b)$

$x = \theta a + b - \theta b$

$x - b = \theta(a - b)$

$\frac{x-b}{a-b} = \frac{b-x}{b-a} = \theta$

$1 - \theta = 1 - \frac{b-x}{b-a} = \frac{x-a}{b-a}$

$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$

(b) Show that

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a} \leq \frac{f(b) - f(x)}{b-x} \tag{2}$$

for all $x \in (a, b)$. Draw a sketch that illustrates this inequality

Solution.

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b) \tag{3}$$

Multiply both sides by $b - a$

$$f(x)(b-a) \leq (b-x)f(a) + (x-a)f(b) \tag{4}$$

$$f(x)(b-a) \leq (b-a)f(a) - (x-a)f(a) + (x-a)f(b) \tag{5}$$

$$(f(x) - f(a))(b-a) \leq -(x-a)f(a) + (x-a)f(b) \tag{6}$$

$$(f(x) - f(a))(b-a) \leq (x-a)(f(b) - f(a)) \tag{7}$$

Divide both sides by $x - a$

$$\frac{(f(x) - f(a))}{x-a}(b-a) \leq (f(b) - f(a)) \tag{8}$$

Divide both sides by $b - a$

$$\frac{(f(x) - f(a))}{x-a} \leq \frac{(f(b) - f(a))}{b-a} \tag{9}$$

Now, let's start with the same inequality, $f(x)(b-a) \leq (b-x)f(a) + (x-a)f(b)$, and work it in a new direction

$$f(x)(b-a) \leq (b-x)f(a) + (x-a)f(b) \tag{10}$$

$$f(x)(b-a) \leq (b-x)f(a) + (b-a)f(b) - (b-x)f(b) \tag{11}$$

$$(f(x) - f(b))(b-a) \leq (b-x)(f(a) - f(b)) \tag{12}$$

$$(f(b) - f(x))(b-a) \geq (b-x)(f(b) - f(a)) \tag{13}$$

Multiply each side by $\frac{1}{(b-x)(b-a)}$

$$\frac{f(b) - f(x)}{b-x} \geq \frac{f(b) - f(a)}{b-a} \tag{14}$$

Put together we get:

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a} \leq \frac{f(b) - f(x)}{b-x} \tag{15}$$

Note: The sketch is on the next page.

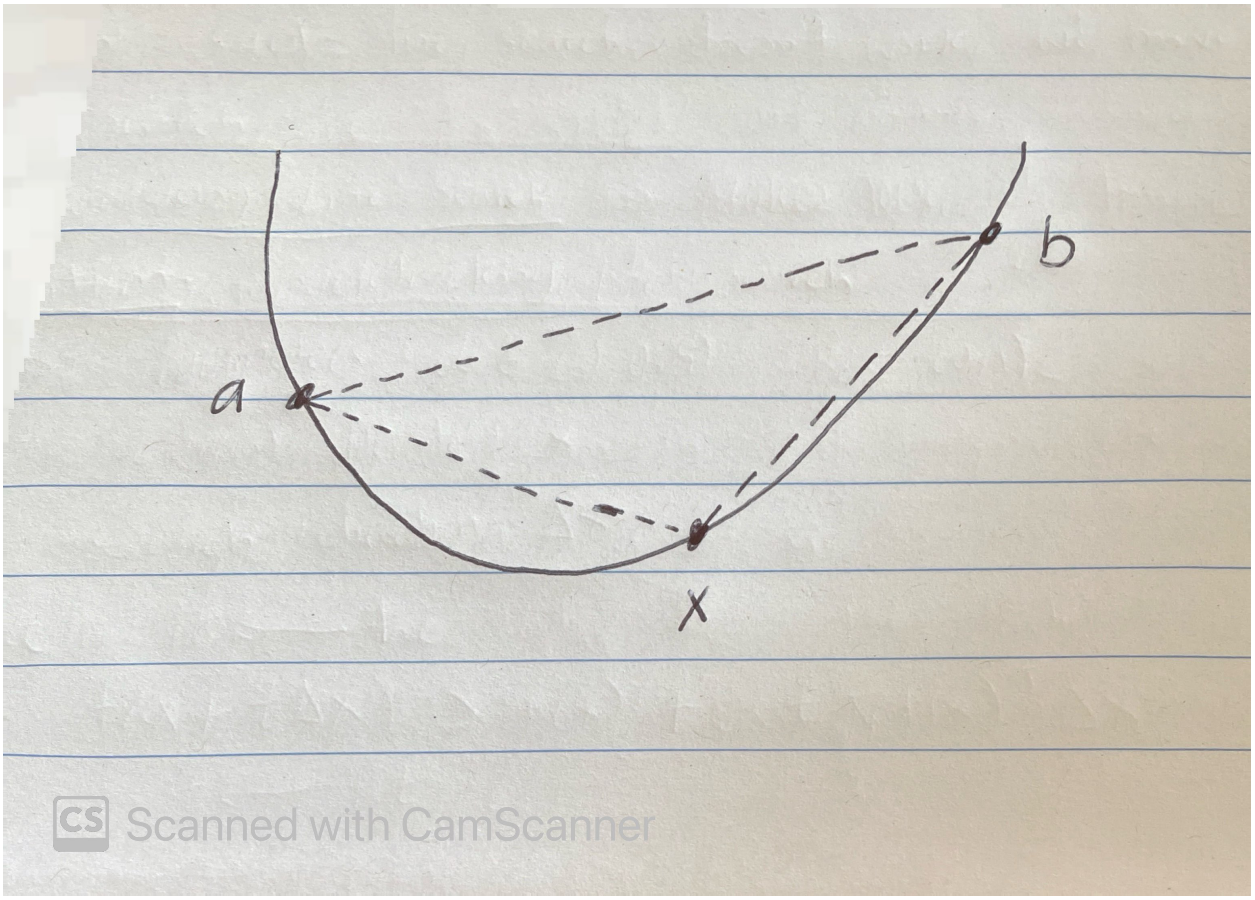

Figure 1: Sketch of Inequality in (B)

Geometrically, this sketch depicts the inequality in Part (B). It shows that the slope of the line connecting points $(a, f(a))$ and $(b, f(b))$ is GREATER than the slope of the line connecting points $(a, f(a))$ and $(x, f(x))$. The sketch also shows that slope of the line connecting points $(a, f(a))$ and $(b, f(b))$ is LESS than the slope of the line connecting points $(x, f(x))$ and $(b, f(b))$

(c) Suppose f is differentiable. Use the result in (b) to show that

$$f'(a) \leq \frac{f(b) - f(a)}{b - a} \leq f'(b) \tag{16}$$

Solution. In Part (b) we proved that $\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x}$

Let's decompose this inequality into its two components:

Component 1: $\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a}$
Component 2: $\frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x}$
If we take the limit as $x \to a$ for Component 1, we see that:

$f'(a) \leq \frac{f(b)-f(a)}{b-a}$

If we take the limit as $x \to b$ for Component 2, we see that:

$\frac{f(b)-f(a)}{b-a} \leq f'(b)$

Putting these results together, we get: $f'(a) \leq \frac{f(b)-f(a)}{b-a} \leq f'(b)$
We can also prove this result via a property of Convex Functions:
We know that, for a convex function, the first-order Taylor approximation is in fact a global underestimator of the function. Hence, we can say the following:

$f(a) + f'(a)(x-a) \leq f(x)$
$f'(a) \leq \frac{f(x)-f(a)}{x-a}$

$f(b) + f'(b)(x-b) \leq f(x)$
$f'(b)(b-x) \geq f(b) - f(x)$
$f'(b) \geq \frac{f(b)-f(x)}{b-x}$
$\frac{f(b)-f(x)}{b-x} \leq f'(b)$

We know that $f'(a) \leq \frac{f(x)-f(a)}{x-a}$, $\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x}$, and $\frac{f(b)-f(x)}{b-x} \leq f'(b)$, we can say that:

$f'(a) \leq \frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x} \leq f'(b)$

$f'(a) \leq \frac{f(b)-f(a)}{b-a} \leq f'(b)$

(d) Suppose f is twice differentiable. Use the result in (c) to show that $f''(a) \geq 0$ and $f''(b) \geq 0$

Solution. The result in Part (c) shows us that $f'(a) \leq \frac{f(b)-f(a)}{b-a} \leq f'(b)$

From this, since $f'(a) \leq f'(b)$, we can go one step further and say that $\frac{f'(b)-f'(a)}{b-a} \geq 0$

From this, we can take the limit and see that:
$f''(a) = \lim_{b \to a} \frac{f'(b)-f'(a)}{b-a} \geq 0$

By extension, we can say that
$f''(b) = \lim_{a \to b} \frac{f'(b)-f'(a)}{b-a} \geq 0$

Problem 2. Show that a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every line segment, its average value on the segment is less than or equal to the average of its values at the endpoints of the segment: For every $x, y \in \mathbb{R}^n$,

$$\int_0^1 f(x + \lambda(y-x))\, d\lambda \leq \frac{f(x)+f(y)}{2} \tag{17}$$

Solution. Step 1: The "Only If" Part
We know that f is convex and that f is continuous. Our goal is to prove that, for every line segment, its average value on the segment is less than or equal to the average of its values at the endpoints of the segment. Let's re-arrange the insides of the integral to write its mathematical equivalent:

$$\int_0^1 f((1-\lambda)x+\lambda y)\,d\lambda \tag{18}$$

Based on the definition of convexity, we know that $f((1-\lambda)x+\lambda y) \leq \lambda f(y)+(1-\lambda)f(x)$

$$\int_0^1 f((1-\lambda)x+\lambda y)\,d\lambda \leq \int_0^1 \lambda f(y)+(1-\lambda)f(x)\,d\lambda \tag{19}$$

Let's evaluate this integral

$$\int_0^1 \lambda f(y)+(1-\lambda)f(x)\,d\lambda = \int_0^1 \lambda f(y)\,d\lambda + \int_0^1 (1-\lambda)f(x)\,d\lambda \tag{20}$$

$$f(y)\int_0^1 \lambda\,d\lambda + f(x)\int_0^1 (1-\lambda)\,d\lambda \tag{21}$$

$$f(y)([\frac{\lambda^2}{2}]_0^1) + f(x)([\lambda - \frac{\lambda^2}{2}]_0^1) \tag{22}$$

$$f(y)(\frac{1}{2}) + f(x)(\frac{1}{2}) = \frac{f(x)+f(y)}{2} \tag{23}$$

We have shown that:

$$\int_0^1 f(x+\lambda(y-x))\,d\lambda \leq \frac{f(x)+f(y)}{2} \tag{24}$$

Step 2: "If" Part
We know that $\int_0^1 f(x+\lambda(y-x))\,d\lambda \leq \frac{f(x)+f(y)}{2}$ and that f is continuous. Our task is to prove that f is convex. Let's approach this proof by proving the contrapositive. Let's say that f is NOT convex. This means that there exists a $\lambda_c \in [0,1]$, for some x and y, where $f((1-\lambda_c)x+\lambda_c y) > \lambda_c f(y)+(1-\lambda_c)f(x)$

Let's define $F(\lambda) = f((1-\lambda)x+\lambda y) - \lambda f(y) - (1-\lambda)f(x)$. Since f is continuous, we know that F is continuous. We also know that, at λ_c, F will have a positive value. Let λ_b be the first point before λ_c where F attains a value of 0. Let λ_d be the first point after λ_c where F attains a value of 0.

We can clearly see that, on the open interval (λ_b, λ_d), $F(\lambda) > 0$. This means that for all λ on the open interval (λ_b, λ_d), we can see that $f((1-\lambda)x+\lambda y) > \lambda f(y)+(1-\lambda)f(x)$

Let $j = (1-\lambda_b)x+\lambda_b y$ and $k = (1-\lambda_d)x+\lambda_d y$
$\int_{\lambda_b}^{\lambda_d} f((1-\lambda)x+\lambda y)\,d\lambda > \int_{\lambda_b}^{\lambda_d} \lambda f(y)+(1-\lambda)f(x)\,d\lambda$
Hence, we can say that, if we make $\theta \in [0,1]$, $\int_0^1 f(j+\theta(k-j))\,d\theta > \int_0^1 \theta f(k)+(1-\theta)f(j)\,d\theta$

Since we know, from our previous work in Step 1, that $\int_0^1 \theta f(k)+(1-\theta)f(j)\,d\theta = \frac{f(j)+f(k)}{2}$

Hence, we can see that $\int_0^1 f(j+\theta(k-j))\,d\theta > \frac{f(j)+f(k)}{2}$

By proving the contrapositive, we have completed the proof

Problem 3. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex with $dom f = \mathbb{R}^n$, and bounded above on $\mathbb{R}^n$. Show that f is constant

Solution. Let's do a proof by contradiction. We will first start with the assumption that f is NOT constant. So there exist α, β on the domain of f such that $f(\alpha) < f(\beta)$. Let $g(\theta) = f(\alpha + \theta(\beta - \alpha))$. We know that f is convex. Furthermore, we learned that convex functions, when restricted to any line in its domain is convex as well. Hence, we can clearly see that g is convex as well. We can also clearly see that $g(0) < g(1)$

Based on the definition of convex functions, we know that:
$g(1) \leq (1 - \frac{1}{\theta})g(0) + \frac{1}{\theta}g(\theta)$.

for values of θ that satisfy $\theta > 1$

Let's rearrange the inequality:

$\frac{1}{\theta}g(\theta) \geq g(1) - (1 - \frac{1}{\theta})g(0)$
$g(\theta) \geq \theta g(1) - (\theta - 1)g(0)$
$g(\theta) \geq g(0) + \theta(g(1) - g(0))$

Since $g(1) > g(0)$, we know that $g(1) - g(0)$ is positive. Hence, the term $g(0) + \theta(g(1) - g(0))$, as θ gets larger and larger, will become larger and larger.

Based on this, as $\theta \to \infty$, we can see that $g(\theta)$ grows without bound. This contradicts the notion that f is bounded above on $\mathbb{R}^n$

Problem 4. *Minimizing a quadratic function.* Consider the problem of minimizing a quadratic function:
minimize $f(x) = \frac{1}{2}x^T Px + q^T x + r$,
where $P \in S^n$ (but we do not assume $P \succcurlyeq 0$)

(a) Show that if P $\not\succcurlyeq 0$, i.e., the objective function f is not convex, then the problem is unbounded below.

Solution. We are given that $P \not\succcurlyeq 0$. This means that P is NOT positive semidefinite.

There exists a z such that $z^T Pz < 0$. Let's say we have a constant $\lambda \in \mathbb{R}^n$ and let's plug in λz into our equation.
$f(\lambda z) = \frac{1}{2}(\lambda z)^T P(\lambda z) + q^T(\lambda z) + r$
$f(\lambda z) = \frac{\lambda^2}{2} z^T Pz + \lambda q^T z + r$

Since the quadratic term dominates and $z^T Pz < 0$, we can see that as λ gets larger, this function's value will get closer and closer to $-\infty$
As $\lambda \to \infty$, we see that $f(\lambda z) \to -\infty$. This means that the functions is UNBOUNDED below

---

(b) Now suppose that P $\succcurlyeq 0$ (so the objective function is convex), but the optimality condition $Px^* = -q$ does not have a solution. Show that the problem is unbounded below.

Since the optimality condition $Px^* = -q$ does not have a solution, q is NOT in the column space of P. Hence, q can be expressed as $q = q_2 + q_3$ where q_2 is the projection of q onto the column space of

P and q_3 is orthogonal to the column space.

Since q_3 is orthogonal to the column space of P, we can say that $q_3^T P q_3 = 0$. Let's have λq_3 for some constant $\lambda \in \mathbb{R}^n$.
$f(\lambda q_3) = \frac{1}{2}(\lambda q_3)^T P(\lambda q_3) + q^T(\lambda q_3) + r$

$f(\lambda q_3) = \frac{\lambda^2}{2} q_3^T P q_3 + \lambda q^T q_3 + r$

We know that $q_3^T P q_3 = 0$. Let's further simplify

$f(\lambda q_3) = \lambda q^T q_3 + r$

We know that $q = q_3 + q_2$. Let's substitute this for q

$f(\lambda q_3) = \lambda(q_3 + q_2)^T q_3 + r$

$f(\lambda q_3) = \lambda q_3^T q_3 + \lambda q_2^T q_3 + r$

We know that q_2 and q_3 are orthogonal. Hence, $q_2^T q_3 = 0$

$f(\lambda q_3) = \lambda q_3^T q_3 + r$

This is unbounded below. As $\lambda \to -\infty$, we see that $f(\lambda q_3) \to -\infty$

Problem 5. *Minimizing a quadratic-over-linear fractional function.* Consider the problem of minimizing the function $f : \mathbb{R}^n \to \mathbb{R}$, defined as:

$$f(x) = \frac{\|\mathbf{Ax} - \mathbf{b}\|_2^2}{c^T x + d}, \quad dom f = \{x | c^T x + d > 0\} \tag{25}$$

We assume $rank A = n$ and $b \notin R(A)$

(a) Show that f is closed.

Solution. Based on the description, it is clear to see that the domain of f is open. Hence, if we prove that for every sequence $x_i \in dom f$ such that $\lim_i x_i = x \in bd\ \ dom f$, we have $\lim_{i \to \infty} f(x_i) = \infty$, then we can show that f is closed.

Since $b \notin R(A)$, we can see that $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ is lower bounded by a positive real number. Hence, for any given sequence H where H is fully contained in the domain of f and the limit point of H is on the boundary of the domain of f, we can see that as H approaches its limit point, $f(x)$ will approach infinity.

Let's further analyze this. Clearly, the boundary of the domain is when $c^T x + d = 0$. Since $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ is lower bounded by a positive real number, we can say that $\|\mathbf{Ax} - \mathbf{b}\|_2^2 \geq k$ where k is some positive real number. So, we can rewrite our function as $f(x) = \frac{\|\mathbf{Ax} - \mathbf{b}\|_2^2}{c^T x + d} \geq \frac{k}{c^T x + d}$. Obviously, as we approach the boundary of the domain, the denominator of $\frac{k}{c^T x + d}$ approaches 0 and $\frac{k}{c^T x + d}$ approaches infinity. Since $f(x) = \frac{\|\mathbf{Ax} - \mathbf{b}\|_2^2}{c^T x + d} \geq \frac{k}{c^T x + d}$, $f(x)$ will approach infinity as well!

(b) Show that the minimizer x^* of f is given by

$$x^* = x_1 + tx_2 \tag{26}$$

where $x_1 = (A^T A)^{-1} A^T b$, $x_2 = (A^T A)^{-1} c$, and $t \in \mathbb{R}$ can be calculated by solving a quadratic equation.

Solution. To find the value of x^* that minimizes f, we know that $\nabla f(x^*) = 0$

$\nabla f(x^*) = \frac{2A^T(Ax^*-b)(c^T x^*+d)}{(c^T x^*+d)^2} - \frac{\|\mathbf{Ax^*-b}\|_2^2(c)}{(c^T x^*+d)^2}$

$\nabla f(x^*) = \frac{2A^T(Ax^*-b)}{(c^T x^*+d)} - \frac{\|\mathbf{Ax^*-b}\|_2^2(c)}{(c^T x^*+d)^2}$

Here is my proposed strategy for the rest of this proof:
To show that $x^* = x_1 + tx_2$ is a valid minimizer, we will substitute the given values of x_1 and x_2 and show that there is a valid value of t such that $\nabla f(x^*) = 0$

$\nabla f(x^*) = \frac{2(A^T A(x_1+tx_2)-A^T b)}{(c^T(x_1+tx_2)+d)} - \frac{\|\mathbf{A(x_1+tx_2)-b}\|_2^2(c)}{(c^T(x_1+tx_2)+d)^2}$

$A^T A x_1 = A^T b$. We can use that to our advantage and substitute it in our expression
$\nabla f(x^*) = \frac{2(A^T b+tA^T Ax_2-A^T b)}{(c^T(x_1+tx_2)+d)} - \frac{\|\mathbf{A(x_1+tx_2)-b}\|_2^2(c)}{(c^T(x_1+tx_2)+d)^2}$

$\nabla f(x^*) = \frac{2tA^T Ax_2}{(c^T(x_1+tx_2)+d)} - \frac{\|\mathbf{A(x_1+tx_2)-b}\|_2^2(c)}{(c^T(x_1+tx_2)+d)^2}$

$A^T A x_2 = c$. We can use that to our advantage and substitute it in our expression

$\nabla f(x^*) = \frac{2tA^T Ax_2}{(c^T(x_1+tx_2)+d)} - \frac{\|\mathbf{A(x_1+tx_2)-b}\|_2^2(A^T Ax_2)}{(c^T(x_1+tx_2)+d)^2}$

$\nabla f(x^*) = A^T Ax_2(\frac{2t}{(c^T(x_1+tx_2)+d)} - \frac{\|\mathbf{A(x_1+tx_2)-b}\|_2^2}{(c^T(x_1+tx_2)+d)^2})$

Now, we must show that, there is a valid value of t that makes this gradient 0.

Now, let's set this gradient to 0

$A^T Ax_2(\frac{2t}{(c^T(x_1+tx_2)+d)} - \frac{\|\mathbf{A(x_1+tx_2)-b}\|_2^2}{(c^T(x_1+tx_2)+d)^2}) = 0$

$\frac{2t}{(c^T(x_1+tx_2)+d)} - \frac{\|\mathbf{A(x_1+tx_2)-b}\|_2^2}{(c^T(x_1+tx_2)+d)^2} = 0$

$t = \frac{\|\mathbf{Ax_1+Atx_2-b}\|_2^2}{2(c^T x_1+tc^T x_2+d)}$

$2t^2 c^T x_2 + 2t(c^T x_1 + d) = t^2\|\mathbf{Ax_2}\|_2^2 + 2t(Ax_1 - b)^T Ax_2 + \|\mathbf{Ax_1 - b}\|_2^2$

We can make some key observations here that allow us to simplify this expression

- $||Ax_2||_2^2 = (Ax_2)^T(Ax_2) = x_2 A^T Ax_2 = (A^T Ax_2)^T x_2 = c^T x_2$
- $2t(Ax_1 - b)^T Ax_2 = 2t(x_1^T A^T - b^T)Ax_2 = 2t(x_1^T A^T Ax_2 - b^T Ax_2) = 2t((A^T Ax_1)^T x_2 - b^T Ax_2) = 2t((A^T b)^T x_2 - b^T Ax_2) = 2t(b^T Ax_2 - b^T Ax_2) = 0$

$2t^2c^Tx_2 + 2t(c^Tx_1 + d) = t^2c^Tx_2 + \|\mathbf{Ax_1} - \mathbf{b}\|_2^2$

$t^2c^Tx_2 + 2t(c^Tx_1 + d) - \|\mathbf{Ax_1} - \mathbf{b}\|_2^2 = 0$

The roots of this equation are:

$t = \frac{-(c^Tx_1+d)\pm\sqrt{(c^Tx_1+d)^2+c^Tx_2\|\mathbf{Ax_1}-\mathbf{b}\|_2^2}}{c^Tx_2}$
As given by the domain restriction, we just need to solve for the value of t such that:

$c^T(x^*) + d > 0$
$c^T(x_1 + tx_2) + d > 0$

Given the equation for the roots, we can proceed from there:

$t = \frac{-(c^Tx_1+d)\pm\sqrt{(c^Tx_1+d)^2+c^Tx_2\|\mathbf{Ax_1}-\mathbf{b}\|_2^2}}{c^Tx_2}$

Multiply both sides by c^Tx_2

$c^Tx_2t = -(c^Tx_1 + d) \pm \sqrt{(c^Tx_1 + d)^2 + c^Tx_2\|\mathbf{Ax_1} - \mathbf{b}\|_2^2}$

$(c^Tx_1 + d) + c^Tx_2t = \pm\sqrt{(c^Tx_1 + d)^2 + c^Tx_2\|\mathbf{Ax_1} - \mathbf{b}\|_2^2}$

$c^T(x_1 + tx_2) + d = \pm\sqrt{(c^Tx_1 + d)^2 + c^Tx_2\|\mathbf{Ax_1} - \mathbf{b}\|_2^2}$
As I showed earlier, $||Ax_2||_2^2 = (Ax_2)^T(Ax_2) = x_2A^TAx_2 = (A^TAx_2)^Tx_2 = c^Tx_2$. Hence, I can rewrite the square root as such:

$c^T(x_1 + tx_2) + d = \pm\sqrt{(c^Tx_1 + d)^2 + ||Ax_2||_2^2\|\mathbf{Ax_1} - \mathbf{b}\|_2^2}$
The inside of the square root, clearly, must be ≥ 0. Hence, the square root will evaluate to a real number! This is good news for us.
Since we want $c^T(x_1+tx_2)+d > 0$, we can just say that we want $c^T(x_1+tx_2)+d = \sqrt{(c^Tx_1 + d)^2 + ||Ax_2||_2^2\|\mathbf{Ax_1} - \mathbf{b}\|_2^2}$

$c^T(x_1 + tx_2) + d = \sqrt{(c^Tx_1 + d)^2 + c^Tx_2\|\mathbf{Ax_1} - \mathbf{b}\|_2^2}$

This means that $t = \frac{-(c^Tx_1+d)+\sqrt{(c^Tx_1+d)^2+c^Tx_2\|\mathbf{Ax_1}-\mathbf{b}\|_2^2}}{c^Tx_2}$ is both guaranteed to be a valid real number that also satisfies our domain constraint!

This completes the proof.
